Novel Rare Genetic Variants Associated with Airflow Obstruction in the General Population

de Vries, Maaike; van der Plaat, Diana A; Nedeljkovic, Ivana; van der Velde, K Joeri; Amin, Najaf; van Duijn, Cornelia M; Vonk, Judith M; Boezen, H Marike; van Diemen, Cleo C

Published in:
American Journal of Respiratory and Critical Care Medicine

DOI:
10.1164/rccm.201909-1868LE

AJRCCM Articles in Press. Published on 18-October-2019 as 10.1164/rccm.201909-1868LE

Novel Rare Genetic Variants Associated with Airflow Obstruction in the General Population

Maaike de Vries, PhD[1,2], Diana A. van der Plaat, PhD [1,2], Ivana Nedeljkovic, MSc[3], K. Joeri van der Velde, PhD[4,5], Najaf Amin, PhD[3], Cornelia M. van Duijn, PhD[3], Judith M. Vonk, PhD[1,2], H. Marike Boezen, PhD[1,2], Cleo C. van Diemen, PhD[4]

1. University of Groningen, University Medical Center Groningen, Department of Epidemiology, Groningen, the Netherlands
2. University of Groningen, University Medical Center Groningen, Groningen Research Institute for Asthma and COPD (GRIAC), Groningen, the Netherlands
3. Department of Epidemiology, Erasmus Medical Center, Rotterdam, the Netherlands
4. University of Groningen, University Medical Center Groningen, Department of Genetics, Groningen, the Netherlands
5. University of Groningen, University Medical Center Groningen, Genomics Coordination Center, Groningen, the Netherlands

Correspondence to:

M. de Vries, PhD
Department of Epidemiology
University Medical Center Groningen, the Netherlands
Hanzeplein 1, 9713 GZ, Groningen, the Netherlands
m.de.vries04@umcg.nl
Tel: 0031 (0)50 361 7033

FUNDING

This work was supported by consortium grant number 4.1.13.007 from the Lung Foundation Netherlands.

ACKNOWLEDGEMENTS

The LifeLines initiative has been made possible by funds from FES (Fonds Economische Structuurversterking), SNN (Samenwerkingsverband Noord Nederland) and REP (Ruimtelijk Economisch Programma).

This research has been conducted using the UK Biobank Resource under application 42273.

Data extraction has been performed using Band, G. and Marchini, J., "*BGEN: a binary file format for imputed genotype and haplotype data*", bioRxiv 308296.

We thank Kate Mc Intyre for editorial assistance.

Supplementary data is available upon request by the corresponding author (m.de.vries04@umcg.nl).

AJRCCM Articles in Press. Published on 18-October-2019 as 10.1164/rccm.201909-1868LE

To the editor:

The development of chronic obstructive pulmonary disease (COPD) is known to be influenced by both genetic and environmental factors (1, 2). Although smoking is generally considered the main environmental risk factor, only 25% of smokers develop COPD (3) while 25–40% of COPD patients have never smoked (4). This indicates that there are large differences in individual susceptibility that may be at least partially attributable to genetic factors. Genome-wide association studies have identified several common genetic variants (minor allele frequency (MAF) >5%) that affect COPD susceptibility, but these only explain a small amount of individual risk to develop COPD (1). Rare genetic variants (MAF <1%) have largely been ignored, since genome wide association studies require huge sample sizes to reach statistical significance for rare variants because of their low allele frequencies in the general population. Rare genetic variants, similar to common variants, may affect COPD either through interaction with or independent of environmental factors such as smoking. In this study we aimed to identify rare genetic variants for COPD independent of environmental factors.

We selected 36 unrelated subjects with airflow obstruction from the general-population based Lifelines cohort study using the following criteria: FEV_1/FVC <70%, FEV_1/FVC <lower limit of normal, never smoked, no environmental smoking, and no occupational exposures to vapors, gases, dust, fumes, and pesticides. We performed whole genome sequencing (WGS) and prioritized predicted pathogenic sequencing variants in exomes using an in-house developed algorithm GAVIN (5). In general, variants present in more than five subjects appeared to be (alignment) artifacts, such as single nucleotide stretches or repeats. Therefore we only included variants that were carried by one to five

subjects. We found 7,355 predicted pathogenic rare variants of which 6,395 variants were carried by only one subject (Figure 1). Of interest, 122 rare genetic variants were carried by 4 or 5 subjects and can actually be considered as common variants in our homogenous identification cohort based on their MAF >5%.

We verified the relevance of these 122 rare variants in COPD in the general population by taking advantage of the imputed data of the UK Biobank. Here, imputed variants with an info-score below 0.5 and genotype values between 0.1 – 0.9 and 1.1 – 1.9 were excluded. Overall, 42 of the 122 variants were imputed. Based on airflow obstruction defined as FEV_1/FVC <70%, participants with imputed genotype data available were divided into cases (n=68,344) and controls (n=374,713). To assess if the frequencies of the 42 imputed rare genetic variants were different between cases and controls, χ^2 tests were performed and a Benjamini-Hochberg adjusted p-value below 25% was considered significant. A total of 7 rare variants present in the genes *SERPINA1, CMYA5, OPA3, SUZ12P1, LRP5, KIF27 and TMC4* were more frequent in individuals with airflow obstruction. While we expected to find only rare variants with large effect sizes, the UK Biobank verification results found effect sizes for rare variants that are, remarkably, not very different from the effect sizes observed for common variants. This further reinforces the idea that very large sample sizes are required for the identification of rare variants in association studies.

By using a fixed effect meta-analysis, we subsequently studied 6 of the 7 variants (the variant in *SUZ12P1* was not imputed) in two smaller general population-based cohorts: Lifelines and the Rotterdam Study. In here, airflow obstruction was defined as FEV_1/FVC <70% and a meta-analysis p-value below 0.05 was considered significant. We replicated the

AJRCCM Articles in Press. Published on 18-October-2019 as 10.1164/rccm.201909-1868LE

three variants present in Kinesin Family Member 27 (*KIF27;* Chr9:86495276), LDL Receptor Related Protein 5 (*LRP5;* Chr11:68174189) and Outer Mitochondrial Membrane Lipid Metabolism Regulator (*OPA3;* Chr19:46088060) (Figure 2).

None of the genes in which the new risk variants are located have been associated with respiratory diseases before. However, based on expression data available within the Genotype-Tissue Expression (GTEx) project, a comprehensive public resource to study tissue-specific gene expression, all three genes are expressed in lung tissue. *KIF27* belongs to the kinesin superfamily of microtubule-based motors. This superfamily plays essential roles in cell division, cell motility, intracellular trafficking, control of microtubule dynamics, and ciliary function (6). *OPA3* has been described as a protein involved in the mitochondrial fission machinery in cells and appears to be well conserved between different cell types (7). *LRP5* is an essential component of the Wnt/β-catenin signalling pathway and known to be important for tissue development and homeostasis. Since all three genes are involved in processes important for the lung, it is highly interesting to further characterize the functional contribution of these variants in COPD.

Although studies on rare genetic variants are currently limited, we are not the first group that endeavored to identify and interpret rare genetic variants associated with COPD. However, compared to initiatives such as targeted re-sequencing and family-based studies (8, 9), our analysis strategy differs in three noteworthy ways: (1) it uses an untargeted, exome-wide approach, (2) it is independent of exposure to cigarette smoke, and (3) our results are applicable to general population cohorts. Nevertheless, we have to acknowledge limitations of our approach. First, we only selected rare genetic variants with a potential pathogenic risk and variants that are not detected by *in silico* prediction tools such as GAVIN

are neglected with our approach. Second, we used imputed data to identify rare genetic variants in the general population. Next to the fact that imputed data is an approximation of the genotype, not all the rare variants could be imputed in all datasets. Finally, the lack of replication of many variants may also be the result of our strategy to replicate only the exact variant in each population. Allele frequencies of rare variants tend to differ between populations and it may be interesting to see whether other variants in the same gene, or a cumulative gene-based risk score, are associated with airflow obstruction.

In conclusion, our novel approach seems promising for exome-wide identification of rare genetic variants and novel candidate genes for COPD that arise independent of exposure to cigarette smoke in the general population.

AJRCCM Articles in Press. Published on 18-October-2019 as 10.1164/rccm.201909-1868LE

REFERENCES

(1) Shrine N, Guyatt AL, Erzurumluoglu AM, Jackson VE, Hobbs BD, Melbourne CA, Batini C, Fawcett KA, Song K, Sakornsakolpat P, Li X, Boxall R, Reeve NF, Obeidat M, Zhao JH, Wielscher M, Understanding Society Scientific Group, Weiss S, Kentistou KA, Cook JP, Sun BB, Zhou J, Hui J, Karrasch S, Imboden M, Harris SE, Marten J, Enroth S, Kerr SM, Surakka I, Vitart V, Lehtimaki T, Allen RJ, Bakke PS, Beaty TH, Bleecker ER, Bosse Y, Brandsma CA, Chen Z, Crapo JD, Danesh J, DeMeo DL, Dudbridge F, Ewert R, Gieger C, Gulsvik A, Hansell AL, Hao K, Hoffman JD, Hokanson JE, Homuth G, Joshi PK, Joubert P, Langenberg C, Li X, Li L, Lin K, Lind L, Locantore N, Luan J, Mahajan A, Maranville JC, Murray A, Nickle DC, Packer R, Parker MM, Paynton ML, Porteous DJ, Prokopenko D, Qiao D, Rawal R, Runz H, Sayers I, Sin DD, Smith BH, Soler Artigas M, Sparrow D, Tal-Singer R, Timmers PRHJ, Van den Berge M, Whittaker JC, Woodruff PG, Yerges-Armstrong LM, Troyanskaya OG, Raitakari OT, Kahonen M, Polasek O, Gyllensten U, Rudan I, Deary IJ, Probst-Hensch NM, Schulz H, James AL, Wilson JF, Stubbe B, Zeggini E, Jarvelin MR, Wareham N, Silverman EK, Hayward C, Morris AP, Butterworth AS, Scott RA, Walters RG, Meyers DA, Cho MH, Strachan DP, Hall IP, Tobin MD, Wain LV. New genetic signals for lung function highlight pathways and chronic obstructive pulmonary disease associations across multiple ancestries. *Nat Genet* 2019;51:481-493.

(2) van der Plaat DA, de Jong K, Lahousse L, Faiz A, Vonk JM, van Diemen CC, Nedeljkovic I, Amin N, Brusselle GG, Hofman A, Brandsma CA, Bosse Y, Sin DD, Nickle DC, van Duijn CM, Postma DS, Boezen HM. Genome-wide association study on the FEV1/FVC ratio in never-smokers identifies HHIP and FAM13A. *J Allergy Clin Immunol* 2016.

(3) Lokke A, Lange P, Scharling H, Fabricius P, Vestbo J. Developing COPD: a 25 year follow up study of the general population. *Thorax* 2006;61:935-939.

(4) Salvi SS, Barnes PJ. Chronic obstructive pulmonary disease in non-smokers. *Lancet* 2009;374:733-743.

(5) van der Velde KJ, de Boer EN, van Diemen CC, Sikkema-Raddatz B, Abbott KM, Knopperts A, Franke L, Sijmons RH, de Koning TJ, Wijmenga C, Sinke RJ, Swertz MA. GAVIN: Gene-Aware Variant INterpretation for medical sequencing. *Genome Biol* 2017;18:6-016-1141-7.

(6) Yue Y, Blasius TL, Zhang S, Jariwala S, Walker B, Grant BJ, Cochran JC, Verhey KJ. Altered chemomechanical coupling causes impaired motility of the kinesin-4 motors KIF27 and KIF7. *J Cell Biol* 2018;217:1319-1334.

(7) Ryu SW, Yoon J, Yim N, Choi K, Choi C. Downregulation of OPA3 is responsible for transforming growth factor-beta-induced mitochondrial elongation and F-actin rearrangement in retinal pigment epithelial ARPE-19 cells. *PLoS One* 2013;8:e63495.

(8) Artigas MS, Wain LV, Shrine N, McKeever TM, UK BiLEVE, Sayers I, Hall IP, Tobin MD. Targeted Sequencing of Lung Function Loci in Chronic Obstructive Pulmonary Disease Cases and Controls. *PLoS One* 2017;12:e0170222.

(9) Qiao D, Lange C, Beaty TH, Crapo JD, Barnes KC, Bamshad M, Hersh CP, Morrow J, Pinto-Plata VM, Marchetti N, Bueno R, Celli BR, Criner GJ, Silverman EK, Cho MH, Lung GO, NHLBI Exome Sequencing Project, COPDGene Investigators. Exome Sequencing Analysis in Severe, Early-Onset Chronic Obstructive Pulmonary Disease. *Am J Respir Crit Care Med* 2016;193:1353-1363.

FIGURE LEGENDS

Figure 1: Study flowchart. Whole-genome sequencing generated 747 million reads per sample on average, yielding an average coverage of 47x per base. Sequence variants were annotated for potential pathogenicity based on gene-specific CADD score thresholds and minor allele frequencies using GAVIN. Alignment artifacts such as single nucleotide stretches or repeats and variants located in mitochondrial DNA, X or Y chromosome were excluded. Distribution of the selected rare variants annotated as (likely) pathogenic among the 36 subjects was assessed. Genetic variants carried by four or five subjects were imputed in UK Biobank (combined Haplotype Reference Consortium and UK10K haplotype resource reference panel) and differences in frequency of the imputed rare genetic variants between cases (FEV_1/FVC <70%) and controls (FEV_1/FVC >70%) were tested using a χ^2 test. By using a fixed-effect meta-analysis, significant variants were replicated in Lifelines (Genome of the Netherlands reference panel) and the Rotterdam Study (Haplotype Reference Consortium reference panel). FEV_1: Forced Expiratory Volume in one second; FVC: Forced Vital Capacity; MAF: Minor Allele Frequency; FDR: False Discovery Rate

Figure 2: Odds ratios of variants that were significantly different between subjects with airflow obstruction and controls in the UK Biobank and meta-analysis of Lifelines and the Rotterdam Study. Odds ratios are presented with confidence intervals. ∞ indicates that no odds ratio could be calculated because this variant was not present in controls.

AJRCCM Articles in Press. Published on 18-October-2019 as 10.1164/rccm.201909-1868LE

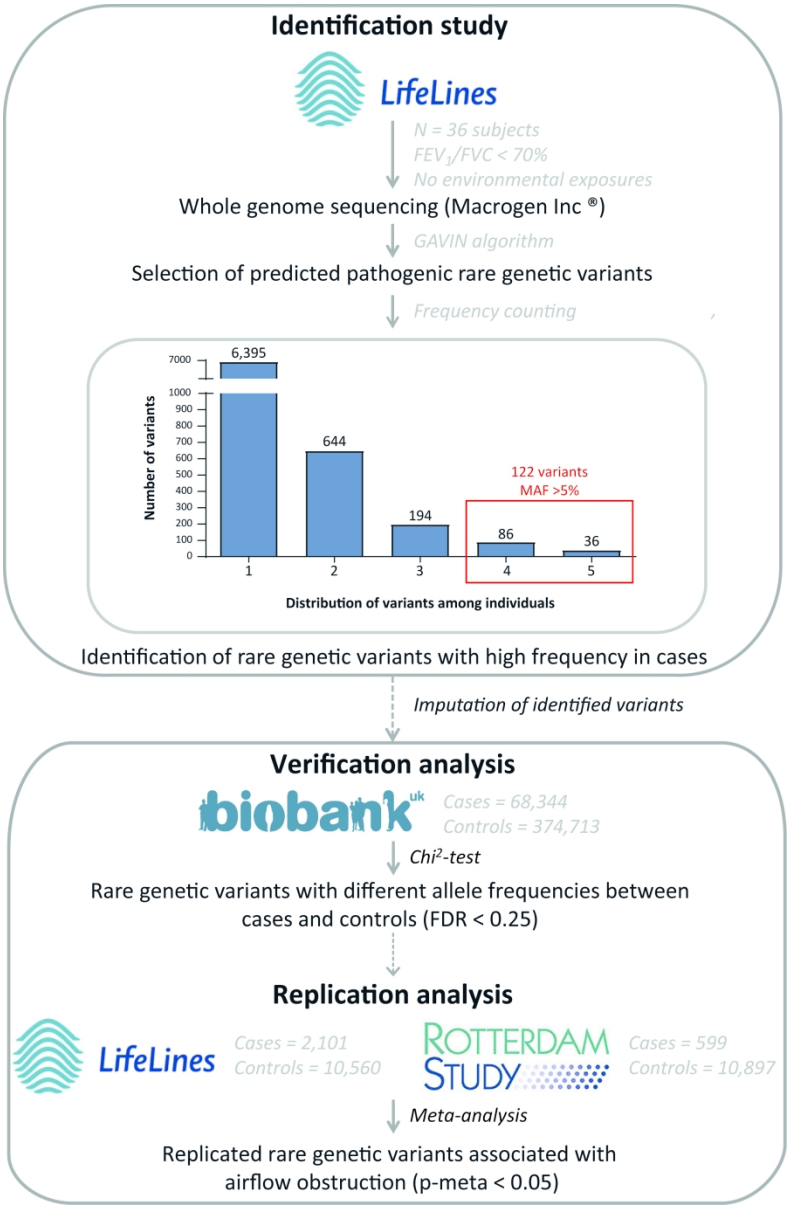

Figure 1: Study flowchart

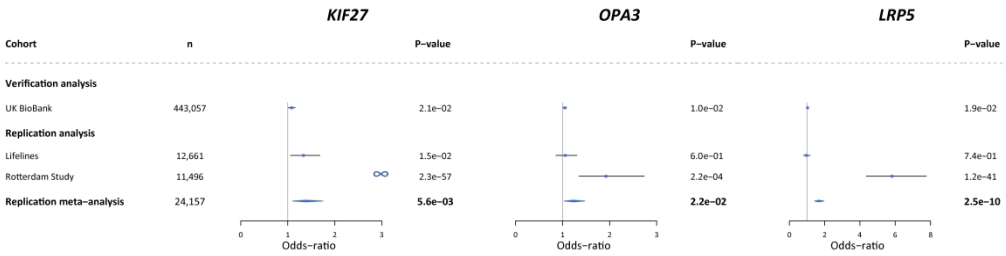

Figure 2: Odds ratios of variants that were significantly different between subjects with airflow obstruction and controls in the UK Biobank and meta-analysis of Lifelines and the Rotterdam Study

354x87mm (300 x 300 DPI)